1967

Comparative Study of Recitation and Review in Seventh Grade Programmed Geography

Jacob Max Setzer
Central Washington University

COMPARATIVE STUDY OF RECITATION AND REVIEW IN SEVENTH GRADE PROGRAMMED GEOGRAPHY

A Thesis

Presented to

the Graduate Faculty

Central Washington State College

In Partial Fulfillment

of the Requirements for the Degree

Master of Education

Jacob Max Setzer

August, 1967

APPROVED FOR THE GRADUATE FACULTY

_____

Otto F. Jakubek, COMMITTEE CHAIRMAN

_____

Alexander H. Howard, Jr.

_____

Daryl Basler

ACKNOWLEDGMENTS

The writer wishes to extend his sincere appreciation to Mr. Otto F. Jakubek for his advice, helpful criticism, and many hours of consultation. Grateful appreciation is also extended to Dr. Daryl Basler for his assistance during the planning and recording of this experiment, and to Dr. A. H. Howard for his generous assistance during the forming of the committee, and for serving on the committee.

And a special note of thanks goes to my wife for her patience and encouragement during the writing of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

THE PROBLEM AND DEFINITIONS OF TERMS USED

Educational research stresses the need for active learning directly involving the student, increased emphasis on the problem-solving approach, and instruction based on individual student requirements. This concept has resulted in considerable experimentation with new methods and materials at the classroom level which has been inspired by a changed relationship between the researcher and the classroom teacher. As seen by Hilgard (16:418), the researcher of learning theory has, in the past, often limited his role to prescriptive assistance. The present attitudes of those engaged in research on learning theory could be described as follows:

> . . . We ought to see whether or not, in the actual context of the classroom, teachers can be helped in their dealing with students to implement these principles with the aid of appropriately designed materials and training in their use, supplemented with tests to see whether or not the desired advances are indeed forthcoming. In other words, we believe that scientific psychology of learning has the obligation to go all the way from theory to practice, using criticized data in every step (16:418).

Another way of summarizing current philosophy is to say that education has four dimensions: the purposes or the "defined behavior changes," the procedures or the "hypothesis about behavior changes," the information or "data and theory about the learner and

learning," and the observations, measurements, and evaluations of the first three dimensions. These four dimensions are essentials of an integrated plan that should be followed in systematic sequence in the classroom (6:391).

An increasing portion of recent classroom experimentation has been concerned with the introduction of programmed material in various subject areas and at various levels. Geography is one of the subject areas in which increasing quantities of programmed material has been made available to teachers. In the initial stages of evaluation, much of the research was preoccupied with comparative studies of programmed materials and conventional methods and materials. The results of such comparisons have usually been summarized as "no significant differences." As Stolurow (8:437) has commented, this is a common though inappropriate pattern in the beginning stages of research on educational innovations. Comparisons at this stage of development have failed to recognize that each of the specific conditions being compared is not the condition, but only one condition selected from a population of possible infinite variation. Therefore, in the initial stages, the parametric study is more useful than the comparative study in determining the relative effectiveness of various programming methods, principles, and conditions. Systematic studies of the latter type will make it possible to identify the conditions essential to successful programmed instruction.

I. THE PROBLEM

Statement of the Problem. The purpose of this study was to compare the effectiveness of two methods of reinforcing learning gained from programmed material in seventh grade geography classes. The means used for evaluating effectiveness were: (1) comparing retention of the factual knowledge of the material by students exposed to the two methods immediately after completing the material, and, after a time lapse of six months, and (2) comparing the ability of students to transfer their factual knowledge of the material immediately after completing the material, and, after a time lapse of six months.

The material used in the experiment consisted of a three book series of programmed instruction in basic geographic skills and concepts published in 1963 by The Macmillan Company under the title of Programmed Geography. Individual books in the series were subtitled as follows: Book I, The Earth in Space; Book II, Continents and Oceans; and Book III, Latitudes and Climates.

The two instructional procedures compared in the study were: (1) the use of programmed material followed by a teacher-directed, systematic review after the completion of each unit; and (2) the use of programmed material followed by an informal recitation after the completion of each unit.

The following specific hypotheses were selected for verification or rejection as a result of findings of the study:

Comparison of student achievement test scores will indicate no statistically significant differences between the systematic review and the informal recitation methods of reinforcing learning gained from programmed geography in the following areas: (1) immediate recall of factual knowledge of the material, and (2) retention of factual knowledge after six months, and (3) immediate ability to transfer factual knowledge of the material, and (4) ability to transfer factual knowledge after six months.

II. IMPORTANCE OF THE STUDY

Experimentation in methods and materials should be followed by evaluation. The advent of programming in the field of geography has resulted in a concurrent need for comparison and evaluation of methods and materials at the classroom level. Schramm (26:50-51) pointed out some of the pertinent areas for research at the present stage of development:

> It seems reasonable at this time to lay aside the question of whether students learn from programmed instruction. They do. So far as we have evidence, they learn a great deal. The useful question is, rather, what are the conditions of effectiveness? From what kinds of programs, trying to teach what kinds of cognitions and behaviors, do students learn most? And for what kinds of teaching, in what situations, with what students, are programs to be preferred to other methods of teaching?

The author's conclusion was that too little was known about either of these two broad questions, and least was known about the second question dealing with kinds of teaching, situations, and students. Any reliable information regarding the latter question would be closely tied to further experimentation regarding the methods used in applying programmed material in the classroom.

Educators in foreign countries as well as in the United States have shown increasing interest in programming as indicated by a series of Canadian studies (25:55). The critique of the studies questioned the value, at the present time, of comparative studies involving programmed instruction and conventional instruction. In fact, unless some kind of teacherless, program-centered educational system was planned, the question of either teachers or programs was never a valid issue. Instead of a program versus conventional method approach, modern research should become more involved with discovering the specific combinations of teacher and programmed instruction that promise the most effective results. Implementation studies of this type should also be concerned with the kinds of schools and classroom organizations best suited to these specific combinations of teacher and programmed instruction.

III. DEFINITIONS OF TERMS USED

Programmed Instruction. This term has been interpreted as meaning any device, machine, or book with the following instructional characteristics: the material to be learned has been arranged into a series of sequential steps leading from familiar concepts to new materials, continual response required from the student, sufficient cues are presented to make successful response likely, and reinforcement given through immediate knowledge of results by the learner.

Conventional Instruction. Throughout the text of this study, the term "conventional instruction" has been interpreted as meaning any instructional method which has a textbook and teacher orientation using a variety of techniques such as assigned readings, lectures, demonstrations, discussions, and projects.

Systematic Review. Systematic review has been interpreted as meaning a teacher-directed, orderly review of the skills and concepts presented in a unit of study.

Informal Recitation. In this study, the term "informal recitation" has been interpreted as meaning an informal question and answer session by students and a teacher after completion of a unit of study.

Research. Research was interpreted as meaning scientific inquiry under laboratory conditions with no immediate goal of application to a specific situation.

Experimentation. The term "experimentation" has frequently been used interchangeably with "action research" and has been interpreted as meaning any investigation which does have an immediate goal of application to a specific situation. This type of inquiry has often been associated with investigations of methods, materials, and teaching techniques at the classroom level.

IV. LIMITATIONS OF THE STUDY

A limiting factor of the study was the grouping of all students, on the basis of reading achievement grades, for the school reading program. As a result of this grouping procedure, the lowest achieving (in reading) half of the seventh grade was by necessity scheduled for one section of geography instruction while the remaining section consisted of the highest achieving (in reading) half of the seventh grade.

V. ORGANIZATION OF REMAINDER OF THE REPORT

The remaining chapters of the report have been organized in the following manner: Chapter II contains a review of the related literature. The methods and procedures used in the study are presented in Chapter III. Chapter IV deals with an analysis of the findings of the study, and Chapter V presents the summer, conclusions, and recommendations for further study.

CHAPTER II

REVIEW OF THE LITERATURE

Few innovations have been the source of greater interest, discussion, and, at times, controversy in educational literature than has the concept of programmed instruction. This discussion has been concerned with the principles, purposes, various applications, and implications of programming. Much of the writing on the various aspects of programming, as well as reports on studies has pertinence and value, but its volume and diversity have made rigorous selection a necessity.

I. DEVELOPMENT OF THE PROGRAM CONCEPT

One of the first applications of programmed learning to modern education has been generally credited to Sidney Pressey. Pressey conducted his experiments at Ohio State in 1926, and his technique employed a teaching machine and a punchboard device. One of the most significant contributions to modern programming of Pressey's approach was that the learner received immediate knowledge of results (8:24-25). Pressey believed that his research had produced highly significant results, but he became disheartened by a lack of public

acceptance of the value of his research. His conclusions were expressed in the following statement written in 1932:

> The writer has found from bitter experience that one person alone can accomplish relatively little, and he is regretfully dropping further work on these problems. But he hopes that enough may have been done to stimulate other workers, that this fascinating field may be developed (9:10).

Pressey's long neglected ideas were given new impetus in 1954 by B. F. Skinner's experiments with programmed learning. Skinner's conception of programmed learning involved the use of a machine which was quite different from Pressey's "teaching machine" in many respects, but was very similar in one basic area. Immediate knowledge of results by the learner was considered just as essential by Skinner as it had been by Pressey. In fact, although there have been many types of programmed instruction and theory, the one common characteristic of programmed instruction has been the method of providing immediate information to the student concerning the results of his response (9:11).

A major difference in the techniques employed by Pressey and Skinner could be best described as a difference in approach or purpose. Pressey's original machine had been designed as a means of providing drill and testing students after they had completed a traditional course of study. Skinner's "teaching machine," on the other hand, was planned to function as a tutor to students who had no prior knowledge on the subject matter presented in the program. Skinner stated:

> This may suggest mass production, but the effect on each student is surprisingly like that of a private tutor . . . (a) there is a constant interchange between program and student . . . the machine induces sustained activity. (b) Like a good tutor the machine insists that a given point be thoroughly understood . . . before the student moves on . . . (c) Like a good tutor the machine presents just that material for which the student is ready . . . (d) Like a skillful tutor, the machine helps the student to come up with the right answer . . . (e) Lastly, of course, the machine, like the private tutor, reinforces the student for every correct response (9:13-14).

II. VARIATIONS IN PROGRAMMING TECHNIQUE

Although there have been many variations in program writing, most of these have stemmed from two basic techniques, notably B. F. Skinner's constructed-response method, and N. A. Crowder's multiple-choice technique.

The best known of the constructed-response programs of the last decade have been those designed by Skinner and his associate Holland. An excellent description of the "Skinnerian" technique can be found in the eight basic rules for programmers suggested by Holland in 1960:

1. Each response must be reinforced immediately.
2. Only overt responses, suitably reinforced, are learned.
3. Errors have an adverse effect on learning.
4. Progress must take place in small successive steps.
5. Aids to the student (cues, prompts) should be withdrawn gradually (the technique called "fading or vanishing").
6. The student's observing behavior should be controlled.
7. Extensive discrimination training is needed to establish an abstraction or concept.
8. The student must write the program (11:48-49).

In the writer's opinion, the three book series of programmed material used in this study conformed to the eight rules suggested by Holland. Also, a Skinner linear technique was utilized.

Although the constructed-response technique has often appeared to dominate the field of programming, Crowder's multiple-choice method has retained active supporters. The essence of Crowder's philosophy was that the student's overt response should not be construed as final and an end in itself, but only as an indication of an internal process. Therefore, the response could be modified, or corrected, and then reinforced. This "branching" technique was basically different from the single track method favored by Skinner (11:61-63).

Stolurow (8:351-353) suggested another possible variation of programmed instruction which has many implications for programmers and educators. His studies offered evidence that individualized instruction might become a reality if sets were sequenced in a manner that encouraged the maximum use of each individual's abilities. Stolurow indicated that with computer-based teaching machines, individual student ability profiles could be stored and used to determine the most effective sequential arrangement of each student's program.

Suppes (28:24-30) reflected general agreement with Stolurow regarding the potential advantages of computer-based programming and described several pilot programs which, in his opinion, illustrated the

practicality of the method. Suppes indicated that due to the tremendous capacity and speed of the computer, numerous students working independently in different places could be accommodated satisfactorily by a single computer.

III. SELECTED RESEARCH REPORTS ON PROGRAMMED INSTRUCTION

A study by Marshall P. Bye (25:27-28) compared the effectiveness of programmed instruction with teacher instruction in seventh and eighth grade mathematics classes. Twenty students in Grade 7 and sixteen students in Grade 8 were matched on the basis of past achievement in mathematics. Thus, each class contained two groups, one of which received programmed instruction while the other received teacher instruction.

Students of Group A worked independently with programmed material under the supervision of a teacher who supplemented the programmed material by teaching a regular lesson whenever it appeared necessary. Students in this group also participated in blackboard drill, and were given spot tests.

Group B was taught by a teacher who followed the program content, but used conventional instructional procedures. Students in Group B were given assignments, exercises, and blackboard drill, but no spot tests were administered. Both groups received individual assistance whenever it was needed.

The experiment required ten class periods of a half hour each, and all work was completed in class. After both groups had finished the material, they were administered an achievement test, and the groups were compared in total and for each grade. The conclusions of the study based on a "t" test analysis of the data were presented as follows:

1. No group did better on programmed instruction than under teacher instruction.
2. The combined classes did as well on programmed instruction as under teacher instruction.
3. The Grade 7 class learned better under teacher instruction than on programmed instruction.
4. The Grade 8 class did as well on programmed instruction as under teacher instruction (25:27).

In other words, the only instance of a statistically significant difference was found between the two seventh grade groups, and this difference favored the teacher-instructed group. However, as stated by the experimenter, the teacher of the program-instructed group had supplemented the programmed material with what could be termed reinforcing activities. Thus, there was a possibility that variables other than the one being tested influenced the results.

A research study by Holland and Porter (9:214-219) involved fourteen graduate students in a programmed educational psychology course at Harvard University. The purpose of the study was to make a comparison of review and repetition of missed items with non-review and non-repetition of missed items. The data of the study was derived from the results of

three tests administered to both groups at appropriate points in the program. In addition to the three tests, a final examination consisting of the same tests was given six months later. An analysis of the data indicated a statistically significant difference in favor of the review group. On the retest, after six months, the review group retained its superiority, but both groups had experienced a significant loss in retention. There was no difference between the two groups in the amount of retention decrement.

Spagnoli (27:447-448) conducted an experiment with ninety sixth grade students of the Grosse Point, Michigan, Public School System. Students from two elementary schools in the system were used in the comparative study of the effectiveness of programmed materials and conventional instructional methods.

Two experimental groups were exposed to programmed books in geography titled "Latitude and Longitude." Conventional methods of instruction were used to teach the same skills to two control groups. The total time of instruction was equal for all groups and lasted two weeks. Each of the two schools involved had one control and one experimental group.

Intelligence and reading tests administered to all the participating students indicated no significant differences within each of the two schools, but very decided differences between schools. In view of these

differences, comparisons were restricted to the control and experimental groups within each school. All of the students received a pretest which had been designed by the teachers of the classes participating in the study. The same test was given as a post-test immediately after the instructional period. A test designed by the publisher of the programmed material was also taken by all students two weeks after the post-test.

Statistical analysis of the data obtained from testing indicated that no significant differences existed between control and experimental groups from pretest to post-test. Analysis of mean scores from the post-test and the "two weeks after" test revealed a similar result. A strong positive correlation was noted between post-test results and the "two weeks after" test of the control groups, but not with the experimental groups. Spagnoli concluded that programmed material was at least as successful as conventional instruction.

A Columbia Basin Research Council study in 1961 by Wriggle and Hite (32:27-32) was concerned with determining how different teacher roles influenced the achievement of students using programmed material in mathematics. Five classes of ninth grade students were used in the experiment. Classes I through IV used a programmed textbook while Class V was taught by a teacher using lesson plans which paralleled the programmed instruction.

The teacher of Class I served only as a monitor and offered

no assistance to students. The teacher of Class II indicated some interest in the material by doing the program, but also gave no help to students. Class III had a teacher who answered any questions asked by students. The teacher of Class IV conducted a review each day after the students completed the program. Class V was not exposed to the programmed material, but was taught by a teacher who used the lecture-recitation approach. Five of the groups described were from Moses Lake, Washington, and a similar grouping of Basin students was arranged. All of the students received pretests and post-tests.

The conclusions of the study, based on statistical evidence, were that students learned as well from a program as they could when taught by a well-prepared teacher. Also, it was found that the kind of contribution of the teacher had a decided effect on student achievement in the programmed classes. One of the group IV classes had a marked superiority in achievement. The teacher of this class had used systematic review as an adjunct to the program. The teacher of the other Class IV had not used review, but had supplemented the program with a question and answer procedure. Students in the class which had received review made greater progress than students from the informal recitation class.

There have been numerous studies made comparing the relative merit of programmed instruction and conventional teaching

methods. In the majority of cases, no clear superiority was given to either procedure. The admitted limitations of many of these experiments had to do with the infinitely possible variations of "conventional teaching methods" resulting from: (1) differences in the specific combinations of readings, lectures, demonstrations, projects, and reinforcing activities encompassed by the term, and (2) differences in teacher philosophy, preparation, and attitudes in regard to the role of the student and the importance of the subject. An additional limitation has to do with the considerable variation possible in programmed material due to differences in mechanical quality, format, and differences in quality of content.

Review of research literature has revealed relatively little research concerned with the most effective role of the teacher in classes using programmed material. The Columbia Basin study by Wriggle and Hite (32:27-31) as previously discussed in this chapter, was the sole example found concerned with the evaluation of various reinforcing activities of the teacher in classes using the same programmed material. The dearth of this type of study in the field of geography was particularly noted.

CHAPTER III

PROCEDURES

I. DESIGN

The study was conducted during the fall of 1966 at Selah, Washington, under the auspices of the administration of School District Number 119. As presented in preceding chapters, the purpose of the experiment was to compare the practice of teacher-directed, systematic review after each unit with informal recitation after each unit. A null hypothesis of no statistically significant differences between control and experimental groups was selected for verification or rejection as a result of findings of the study.

Students who participated were seventh graders from the junior high school, and the experiment was part of their regular course work in the geography of the Eastern Hemisphere. The programmed material selected as the basis for the study had been used for two years to provide students with the basic skills and concepts necessary to later course study of the Eastern Hemisphere. Completion of the experiment required ten weeks and three days of class time, from September 1st to November 14th. Three days were needed to administer a group intelligence test, and the two standardized tests of the study during the fall

quarter. The latter two tests, which had been selected for evaluation of the study, were given again six months later on May 16th and 17th. This additional testing required another two days of class time.

There were a total of 156 students in the seventh grade, and they had been divided into a morning section an an afternoon section for geography instruction. This had been done by the school administration prior to the beginning of the school year 1966-67. The two sections were approximately even in number, but not in potential achievement. This inequality resulted from the required grouping of all students for the school's reading program.

It should be noted that the grouping of students for the reading program had been accomplished primarily on the basis of reading grades with a second criterion of reading grade levels. The reading grades had been arranged from the highest to the lowest, and the students were then divided into eight reading classes. The reading grade levels of the students were used as an added check on accurately placing individuals in a reading class composed of students of similar potential ability. As a result of the grouping procedure for reading, the lowest four reading classes became the morning section of geography while the highest four classes became the afternoon section for geography instruction. There were some discrepancies to this grouping arrangement caused by student transfers in and out of the district, and from students being changed from

one geography section to the other, but the above description applied to the majority of students.

The Lorge-Thorndike group intelligence test was administered to both sections of seventh grade geography on September 1st for the purpose of matching pairs of students, one from a control group and one from the experimental group of the same section, on the basis of intelligence quotients. Examination of test results after the scores had been recorded indicated there was some overlapping of intelligence test scores between students in the two sections of geography. A statistical analysis of the scores of the two sections revealed that the mean intelligence quotient of Section I (morning group) was 100, and the standard deviation 10.8. In contrast to this, the mean intelligence quotient of Section II (afternoon group) was 118, and the standard deviation 10.6. It should also be added that the standard error for level four of the Lorge-Thorndike Verbal Battery has been given as 4.6 in the technical manual for the test. This was the test administered on September 1st and subsequently used as a basis for matching students.

A considerable number of the original 156 students in the experiment were eliminated during the course of the school year for various reasons such as: transfer out of the district, extended absence during the experiment and initial testing period, and absence from school on May 16th and 17th when the two tests used for evaluation were

administered for the second time. In addition, nine special education students attended geography classes sporadically and could not be included in the study. Therefore, because of the above reasons, and the fact that the loss of one member of a matched pair usually meant the loss of both members, only 122 of the initial 156 students completed the experiment.

Two of the three teachers participating in the experiment met daily with the control and experimental classes in two rooms located on either side of a reference center. Class meetings were for fifty minutes each day, and there were no contacts between the students in the two classes during this time. All three of the instructors had used the programmed material the previous year, and were accustomed to working together in a team teaching situation. A schedule was adopted which provided for the daily rotation of teachers in the control and experimental classes, and also provided for a planning period every third day for each instructor. Thus, on any given day, one teacher met with the morning and afternoon control groups, and a second teacher met with the morning and afternoon experimental groups. The third instructor was scheduled for two planning periods. This was alternated so that each teacher had two consecutive days of class duty followed by a day of no class responsibilities. The schedule worked out well in practice since on the few occasions that an instructor was absent, on a day during which he had class responsibilities, the extra teacher was available and

could be easily substituted.

Since it did not appear that any real purpose would be served by discussing the experiment with the students involved in the study, they were not informed. They became accustomed to changes of teachers, and changes in grouping for different class activities as a result of the team organization of the social studies department, and remained unaware that they were participants in an experiment. The seemingly undue amount of testing was also accepted by the seventh graders with complete equanimity.

II. MATERIALS AND METHODS

The materials used by all students in the study consisted of a three book series by Sullivan Associates titled Programmed Geography, and published by The Macmillan Company in 1963. Each book was subtitled and divided into units of study as follows: Book I, The Earth in Space contained eleven units; Book II, Continents and Oceans was divided into nine units; and Book III, Latitudes and Climates also contained eleven units. Thus, the complete series contained a total of thirty-one units of study. Although there was some variation in the length of units, most students completed a unit in from one to two days of class time.

Students worked independently with the programmed material under the supervision of a teacher, and all work was completed in class.

The programmed books remained in the rooms at all times, and teachers in both the control and experimental classes adopted a uniform practice of offering assistance and encouragement to any students having difficulty.

Optimum conditions of unlimited time per unit with each student working at his own pace could not be permitted because of scheduling pressures. Therefore, a reasonable compromise had to be found that was fair, acceptable, and would not do injury to the purpose of the study. Staff members finally decided that whenever eighty per cent of the students in a class had finished a unit, their instructor would allow the remaining students an additional ten minutes to finish the unit. This was usually sufficient to allow most students to complete the unit. However, the remaining few who had not finished were requested to come back to class during their study hall period and complete their assignments. The students concerned appeared to consider this reasonable and just since they realized there were no other assignments of any kind. The above policy regarding time allotments for units was carried out throughout the experiment. Successful implementation of the policy depended on close supervision of classes by teachers in order to be certain that all students kept up to date with assignments. This was a time consuming task for teachers, but one worthwhile benefit was the fact that any given student either finished his daily assignment in class, or before the next class meeting. The policy worked well in practice and no serious problems were encountered.

III. EQUATING THE GROUPS

Section I, the morning class, was divided into two units in which students were randomly placed. These two units served as a control and experimental group respectively. The students of the control and experimental groups were then equated by matching pairs on the basis of results from the Lorge-Thorndike (Verbal Battery) group intelligence test. This had been administered to students at the beginning of the school year, September 1st, through the cooperation of the school administration.

Section II, the afternoon class, was also divided into two units, and a procedure identical to the one used for Section I was followed in equating students of the control and experimental groups of Section II. Completion of these procedures resulted in the formation of a control and experimental group for Section I, and a control and experimental group for Section II.

Students in the control and experimental groups were coded for identification purposes. For example, a given student from a control class would be given the letter C and a numeral, and his twin in the experimental group of the same section would be given the letter E and the same numeral.

Tables III and IV have been placed in the appendix as an illustration of the matching procedure followed in the study. Table III presented the intelligence test scores of students in the experimental

and control groups of Section I, and Table IV provided identical information concerning students in the experimental and control groups of Section II. As may be noted, the majority of the students in both sections were matched within one or two points, although in some cases three points were required to match a pair. The two tables also illustrated the following factors: the overlapping of intelligence test scores between Section I and Section II, the range of scores from 76 to 121 in Section I, and the range of scores from 96 to 139 in Section II.

IV. EXPERIMENTAL GROUPS

The instructional period of the study lasted for ten weeks. During this period, the two experimental groups were exposed to three programmed books dealing with basic geographic skills and concepts. Classes were supervised by a teacher who was available to answer questions, and also provided assistance and encouragement for students. After completion of each unit, the students were given a teacher-directed systematic review of the skills and concepts presented in the programmed material. These reviews were pre-planned by the teacher in an effort to be thorough. After completion of the review, the students were assigned the following unit and instructed to work independently.

V. CONTROL GROUPS

The instructional period of the two control groups was of the

same length as for the experimental groups. The same materials were used, and the groups were also supervised by a teacher who answered questions, and provided assistance and encouragement for students. After each unit, the two control groups participated in an informal recitation period consisting of questions and answers by students and the instructor. There was no preplanned structure designed by the teacher, nor was any attempt made to direct discussion toward comprehensive coverage of the unit.

VI. SOURCE OF DATA AND EVALUATION

Evaluation of the experiment was made on the basis of results from two tests which were administered twice to all students from the four groups in the study. The tests were given the first time on November 15th and 16th, immediately after the four groups had completed the three programmed books. They were administered for the second time on May 16th and 17th. Thus, there was a time interval of approximately six months between the two administrations.

Since the same two tests were given twice, the question of the possible influence of a "practice effect" had to be considered. In an endeavor to secure authoritative opinion concerning this question, a written request for information regarding pertinent studies was made to the directors of the two organizations who supplied the two standardized tests selected for evaluation of the study. An additional request

was made for a personal opinion regarding "practice" effect, in reference to their test, in the event they had no knowledge of research evidence. In answer to the latter request, Cynthia D. Buchanan (4), Director of Programming, Sullivan Associates, replied as follows:

> . . . there will be no effect on the test scores resulting from learning of our geography test. You are thus safe in using the tests as a measure of differences in retention.

In answer to the same question, D. A. N. Hieronymus (15), Director, Iowa Basic Skills Testing Program, replied in the following manner:

> It might possibly result in inflated gains, but I would not expect much, if any, effect. And whatever effect there would be would apply to both methods groups.

Neither of the two directors had knowledge of research on the "practice" effect in reference to their specific tests.

The tests administered to the students on November 15th was a standardized final test on the three books constructed by the authors of the programmed series. There were 63 questions in the test, but many questions required several answers. For tabulation purposes, each answer was given a point value of one with the exception of the last answer on the test. The latter required two paragraphs of reading and interpretation, and for this reason was given a value of four points. On this basis, the test had a total value of 160 points. The test was designed as a comprehensive evaluation of a student's ability to recall the factual knowledge of the programmed series. The purpose of giving the same

test again six months later was to measure possible differences in retention of factual knowledge between control and experimental groups. For recording purposes, the November 15th administration was designated as test 1A, and the May 16th administration as test 2A.

The other test used in the study was selected to measure differences in the ability of students to transfer their factual knowledge of the programmed material. The test selected for this purpose was the Iowa Test of Basic Skills, Form 1, Test W-1 for the seventh grade. There were 41 multiple-choice questions in this test, and each question was given a point value of one. The same test was given again six months later to measure possible differences in students' ability to transfer their factual knowledge of the programmed series. The November 16th administration of the test was designated as 1B, and the May 17th administration as 2B.

All test papers were corrected immediately and stored until the end of the school year, and, at this time, a t-test was used for analysis of the data resulting from the four tests.

CHAPTER IV

ANALYSIS OF DATA

This experiment was conducted for the purpose of comparing two of the many possible reinforcing procedures which could be utilized by a teacher in classes working with programmed geography. The two reinforcing methods compared in the study consisted of a teacher-directed review after the completion of each unit in the experimental classes, and an informal recitation after the completion of each unit of study in the control classes.

The material used in the study consisted of a three book series by Sullivan Associates, titled Programmed Geography, which had been published by The Macmillan Company in 1963. All of the 156 students in the seventh grade completed the three book series during a ten-week period in the autumn of 1966. However, as previously noted, only 122 of these 156 could be utilized for experimental purposes.

Evaluation of the study was accomplished on the basis of two tests which were administered twice to all subjects in the experiment. The test given to students on November 15th, and again on May 16th, was a final test on the series published by the authors of the three books.

This test was selected as a measurement of immediate recall of program information, and of retention of the material after six months. The total point value of the test was set at 160 points. For recording purposes, the November 15th administration of the test was designated as 1A, and the May 16th repetition of the test as 2A.

The test given to students in the experiment on November 16, and again on May 17th, was selected as a measurement of immediate transfer of program material, and transfer of program information after six months. This test was the Iowa Test of Basic Skills, Form I, Test W-1, and consisted of 41 multiple-choice questions. Each question was given a point value of one point. In order to simplify the recording of grades, the test was designated as 1B on the November 16th administration date, and as 2B for the May 17th repetition of the test.

At the end of the school year 1966-67, the collected data were analyzed through the application of the t-test for the purpose of determining whether statistically significant differences existed between the experimental and control groups. The formula used for the t-test in this study corresponded with the formula by Guilford (12:184) as illustrated below:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{\Sigma x_1^2 + \Sigma x_2^2}{Ni\ (Ni - 1)}}}$$

All results from the t-test analysis of data in this study have been reported at the .05 level of significance. A "t" value of 2.00 or greater was required for evidence of statistically significant differences between the mean scores of the control and experimental groups in the study.

Table I presented the mean scores and "t" scores of the experimental and control groups of Section I (62 students). As may be observed, there were no statistically significant differences between the experimental and control groups of Section I for any of the four tests. As evidenced by the "t" scores of this section, there were no differences of statistical significance in the following areas: immediate recall of program information (test 1A), retention of program material after six months (test 2A), immediate ability to transfer program learnings (test 1B), and ability to transfer program material after six months (test 2B).

TABLE I

MEAN SCORE COMPARISONS OF SECTION I

Test	Experimental Mean	Control Mean	df	t
1A	137.97	132.23	60	1.267
2A	133.35	133.58	60	.049
1B	14.97	15.19	60	.46
2B	18.45	18.03	60	.293

Table II presented the mean scores, and "t" scores of the experimental and control groups of Section II (60 students). As may be observed, there were no statistically significant differences between the experimental and control groups of Section II for three of the four tests. As indicated by the "t" scores of Section II, there were no differences of statistical significance in the following three areas: retention of program material after six months (test 2A), immediate ability to transfer program learnings (test 1B), and ability to transfer program material after six months (test 2B). However, there was a statistically significant difference between the experimental group and control group of Section II in test 1A (immediate recall). As indicated in Table II, the "t" score for test 1A is 2.222.

TABLE II

MEAN SCORE COMPARISONS OF SECTION II

Test	Experimental Mean	Control Mean	df	t
1A	149.37	142.83	58	* 2.222
2A	147.57	146.47	58	.001
1B	23.77	22.83	58	.743
2B	26.13	25.76	58	.347

*Significant at the .05 level

The findings of the study based on the analysis of data presented in this chapter are discussed in Chapter V. A summary of the study

and recommendations for further research are also presented in Chapter V.

CHAPTER V

SUMMARY AND CONCLUSIONS

I. SUMMARY

The purpose of this study was to compare the effectiveness of two methods of reinforcing learning gained from programmed material in seventh grade geography classes. The two methods compared in the experiment consisted of: (1) the use of programmed material followed by a teacher-directed, systematic review after each unit of study, and (2) the use of the same programmed material followed by an informal recitation session by students and a teacher after completion of each unit. The specific hypotheses selected for verification or rejection as a result of findings of the study were:

> Comparison of student achievement test scores will indicate no statistically significant differences between the systematic review and the informal recitation methods of reinforcing learning gained from programmed geography in the following areas: (1) immediate recall of factual knowledge of the material, and (2) retention of factual knowledge after six months, and (3) immediate ability to transfer factual knowledge of the material, and (4) ability to transfer factual knowledge after six months.

The experiment was conducted during a ten-week period in the fall of 1966 through the cooperation of the administration of School District Number 119 at Selah, Washington. Three social studies

teachers, working together in a team organization, supervised and directed the 122 seventh grade students participating in the study. An effort was made to minimize possible teacher variables by scheduling the three instructors in a manner that allowed each an equal time in the two control and two experimental classes.

The seventh grade students had been divided into two sections for geography instruction by the school administration prior to the beginning of the school year 1966-67. To accomplish the purpose of the study, each of the two sections was divided into two units in which students were then randomly placed. Subjects in the two units were matched on the basis of results from the Lorge-Thorndike group intelligence test which had been administered on September 1st. One unit of each section was labeled control and the other designated as experimental. Thus, each of the two sections had a control and an experimental group. Section I (morning group) contained 62 of the 122 students who completed the experiment, and Section II (afternoon group) contained the remaining 60 students.

The material used in the experiment consisted of a three-book series of programmed material titled Programmed Geography, published by The Macmillan Company. The three books contained a total of 31 units. Students in both the control and experimental groups worked independently with these materials in separate rooms. Each room was supervised by a teacher who provided any needed assistance. After completion of a

given unit, the experimental groups were given a teacher-directed, orderly review of the unit, and the control groups participated in an informal recitation session.

Evaluation of the study was based on results from two tests which were administered once immediately after the two control and two experimental groups had completed the programmed material, and for a second time six months later. Test A was selected for evaluation of students' ability to recall factual knowledge gained from the programmed material. This test was a comprehensive final test included in the three book series by Sullivan Associates, the authors of Programmed Geography. In order to facilitate recording, the first administration of the final test was designated 1A, and the second 2A. The other test used was the Iowa Test of Basic Skills, Form 1, Test W-1, for the seventh grade. This test was selected to measure the ability of students to transfer their factual knowledge of the programmed series. For recording purposes, Test W-1 was designated as 1B for the first administration, and as 2B when given for the second time.

The data were analyzed at the conclusion of the school year 1966-67 and results have been presented in Table I on page 31 and Table II on page 32. As indicated in the two tables, a t-test was used for comparison of the mean scores of the experimental and control groups. It should be noted that a "t" of 2.00 or greater was required, at the .05 level of

significance, for evidence of statistically significant difference between control and experimental group mean scores.

II. CONCLUSIONS

In considering the analysis of data for this study, one factor is of prime importance and must be considered for a valid conclusion. This factor is the marked difference between Section I and Section II students in potential school achievement on the basis of reading grades and intelligence test scores. As discussed in Chapter III, the top four reading classes in the seventh grade, from the standpoint of reading grades and reading grade levels, had been placed in Section II geography, while the lowest four reading classes had been placed in Section I. In addition, the mean intelligence score of Section II was 118, and the standard deviation was 10.6, while the mean intelligence score of Section I was 100, and the standard deviation 10.8. Therefore, in view of the considerable differences existing between the two sections, no cross comparisons between Section I and Section II were attempted in this study.

An examination of Table I, from this viewpoint, leads to the conclusion that there was no apparent advantage for either the review or recitation method of reinforcing learning gained from programmed geography in the following areas: immediate recall, recall after six

months, immediate transfer, and ability to transfer after six months. Therefore, the four specific null hypotheses may be accepted.

In reference to Table II, a similar conclusion could be reached with one exception. As indicated by the "t" of 2.222 for test 1A (immediate recall), there was a statistically significant difference in favor of the experimental (review) group. Thus, there was an apparent advantage in the review method of reinforcing programmed learning in the area of immediate recall. Therefore, the null hypothesis of no statistically significant differences between the review and recitation methods in the area of (1) immediate recall must be rejected. As may be noted upon examination of the results of the remaining three tests in Table II, there were no other evidences of superiority for either the review or recitation methods of reinforcing programmed geography instruction. In view of this, null hypotheses number (2) recall after six months, number (3) immediate transfer, and number (4) transfer after six months, may be accepted.

In summary, since the advantage of the review group of Section II in the area of immediate recall was the only exception to full acceptance of the four specific null hypotheses, this single area of superiority for the review method would appear to be of questionable value; particularly so, since there was no evidence of superior recall ability after six months by the review group of Section II. The "t" of .001 resulting from analysis of test 2A supports the latter conclusion.

III. RECOMMENDATIONS

Recommendations for further desirable research in this area would include a similar study with the addition of a third group. As in this study, learning from programmed material would be reinforced by teacher-directed review after each unit in one class, and by informal recitation in the second class. The third class would not participate in any reinforcing activity after a unit, but would merely proceed on to the next unit. A testing program similar to the one used in this study could be utilized because of the desirability of data concerning immediate recall and transfer, and the ability to recall and transfer program learning after a period of time. It is intriguing to speculate on the outcome of a study in which two groups received reinforcement after units and a third group received none.

It would also be desirable to know more about the relative effectiveness of various reinforcing techniques with low, middle, and high ability groups. For example, it would be useful to know if review after programmed geography instruction would affect the three groups equally, differently, or not at all. There are numerous other possibilities involving ability groups and methods of implementing and reinforcing programmed instruction in geography classes. However, the use of three ability groups and several techniques would require a fairly large number of students. Studies of the latter type might best be attempted in a large school, or perhaps several schools could be used.

BIBLIOGRAPHY

SELECTED BIBLIOGRAPHY

1. Arnstine, Donald G., "The Language and Values of Programmed Instruction," Part One, The Educational Forum, 28:219-226, January, 1964.

2. _____, "The Language and Values of Programmed Instruction," Part Two, The Educational Forum, 28:337-346, March, 1964.

3. Brethower, Dale M., Programmed Instruction, A Manual of Programing Techniques, Chicago: Educational Methods, Inc., 1963. 268 pp.

4. Buchanan, Cynthia D., Personal Correspondence, May 26, 1967.

5. _____, (Program Director, Sullivan Associates), Programmed Geography, Books I, II, III. New York: The Macmillan Company, 1963.

6. Colardarci, Arthur P., "The Relevance of Psychology Education," Foundations of Education, George F. Kneller, (ed.). New York: John Wiley and Sons, Inc., 1963. pp. 380-402.

7. Cronbach, Lee J., Educational Psychology, Second Edition, New York: Harcourt, Brace and World, Inc., 1963. 706 pp.

8. DeCecco, John P. (ed.), Educational Technology, Readings in Programmed Instruction, New York: Holt, Rinehart and Winston, 1964. 479 pp.

9. Deterline, William A., An Introduction to Programed Instruction, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1962. 131 pp.

10. Edling, Jack V., et al, Four Case Studies of Programed Instruction, New York: Fund for the Advancement of Education, June, 1964. 115 pp.

11. Fry, Edward B., Teaching Machines and Programmed Instruction, An Introduction, New York: McGraw-Hill Book Company, Inc., 1963. 244 pp.

12. Guilford, J. P., Fundamental Statistics in Psychology and Education, Fourth Edition, New York, New York: McGraw-Hill Book Co., 1965. 606 pp.

13. Hanna, Paul R., Rose E. Sabaroff, Gordon Davies, and Charles R. Farrar, Geography in the Teaching of Social Studies, Concepts and Skills, Boston: Houghton Mifflin and Company, 1966. 511 pp.

14. Hanson, Lincoln F. (ed.), Programs, '63, Compiled and Produced by the Research Division, The Center for Programed Instructions, Inc., in cooperation with the U. S. Department of Health, Education, and Welfare, Office of Education, Washington, D. C., U. S. Government Printing Office, 1963. 814 pp.

15. Hieronymus, A. N., Personal Correspondence, May 11, 1967.

16. Hilgard, Ernest R. (ed.), Theories of Learning and Instruction, The Sixty-third Yearbook of the National Society for the Study of Education, Part I., Chicago: The University of Chicago Press, 1964. 430 pp.

17. Lindquist, E. F., and A. N. Hieronymus, Iowa Test of Basic Skills, Multi-level Edition for Grades 3-9, Form I, Boston: Houghton Mifflin Co., 1955.

18. Lorge, Irving, and Robert L. Thorndike, The Lorge Thorndike Intelligence Tests, Technical Manual, Revised Edition, Boston: Houghton Mifflin Co., 1962.

19. Margulies, Stuart, and Lewis D. Eigen, Applied Programed Instruction, New York: John Wiley and Sons, Inc., 1962. 387 pp.

20. Markle, Susan Meyer, A Programed Primer on Programing, Volume II, New York: The Center for Programed Instruction, Inc., 1961.

21. _____, and Richard M. Bossone, "Programed Materials: A Teaching Aid," The Clearing House, 38:148-151, November, 1963.

22. _____, Lewis D. Eigen, and P. Kenneth Komoski, A Programed Primer on Programing, Volume I, New York: The Center for Programed Instruction, Inc., 1961.

23. Mones, Leon, "Programed Learning, Automation of Teaching and Learning," The Clearing House, 38:136-143, November, 1963.

24. Ofiesh, Gabriel D., and Wesley C. Meierhenry, Trends in Programmed Instruction, Department of Audiovisual Instruction, National Education Association, 1964. 289 pp.

25. Research Division, Canadian Teachers' Federation, Teacher Research on Programmed Instruction, A Collection and Critique of Studies, Ottawa, Ontario: Research Division, Canadian Teachers' Federation, C-64206, April, 1965.

26. Schramm, Wilbur, Programed Instruction Today and Tomorrow, New York: The Fund for the Advancement of Education, November, 1962. 74 pp.

27. Spagnoli, Joseph, "An Experience with Programmed Materials," The Journal of Educational Research, 58:447-448.

28. Suppes, Patrick, "Plug-In Instruction," Saturday Review, 49:25-30, July 23, 1966.

29. The Use of Programed Instruction in U. S. Schools, Report of a Survey of the Use of Programed Instructional Materials in the Public Schools of the United States During the Year 1961-1962, Number OE-34022, Research Division of the Center for Programed Instruction, Inc., in cooperation with the U. S. Department of Health, Education, and Welfare, Office of Education, Washington, D. C., U. S. Government Printing Office, 1962, 83 pp.

30. Tracey, William R., "Programing-Evaluating Achievements," The Clearing House, 38:143-148, November, 1963.

31. Varty, Jonathan W., "Programmed Instruction for Prospective Teachers," The Educational Forum, 28:227-235, January, 1964.

32. Wriggle, Larry, and Dr. Herbert Hite, The Amount and Nature of Teacher Help Necessary for Optimum Achievement Through Use of Programed Learning Devices, Research Report Number 05-01, Olympia, Washington: Office of Superintendent of Public Instruction, December, 1962.

APPENDIX

TABLE III

SCORES OF SECTION I STUDENTS ON THE LORGE-THORNDIKE INTELLIGENCE TEST

Student Number	Experimental Group (E)	Control Group (C)
1	118	121
2	111	114
3	112	115
4	113	112
5	113	112
6	112	111
7	110	110
8	109	110
9	108	108
10	106	107
11	105	106
12	105	105
13	103	104
14	103	103
15	102	102
16	101	101
17	99	99
18	99	98
19	97	98
20	96	96
21	96	96
22	96	96
23	95	95
24	94	94
25	92	92
26	90	90
27	87	90
28	83	82
29	81	81
30	79	79
31	76	77

TABLE IV

SCORES OF SECTION II STUDENTS ON THE LORGE-THORNDIKE INTELLIGENCE TEST

Student Number	Experimental Group (E)	Control Group (C)
1	139	139
2	135	135
3	131	133
4	130	131
5	129	129
6	128	128
7	127	127
8	127	127
9	127	126
10	123	124
11	123	123
12	122	121
13	121	121
14	119	119
15	119	118
16	118	118
17	118	117
18	116	116
19	115	112
20	112	112
21	112	111
22	110	110
23	110	109
24	109	109
25	108	108
26	107	107
27	104	104
28	102	104
29	100	100
30	99	96